6.10

OSSE DATA FLOW AND ANALYSIS

Mark Strickman, W. Neil Johnson, Robert Kinzer,
James Kurfess, Mark Leising
Naval Research Laboratory, Washington D.C.

Robert Cameron, Gregory Jung
University Space Research Association/NRL

Craig Jensen
Applied Research Corporation/NRL

David Grabelsky, Steven Matz, James Pendleton,
William Purcell, Melville Ulmer
Northwestern University, Evanston Ill.

N00173-85-C-2501
1990

I. Introduction

The Oriented Scintillation Spectrometer Experiment (OSSE) consists of four large volume NaI detectors that are optimized to observe gamma-rays in the 0.05 - 10.0 MeV energy band. Each detector can be oriented independently and has a wide variety of observation and telemetry modes, leading to a complex data analysis environment.

The OSSE data analysis task is rendered even more complicated by the on-orbit environment, which exhibits a rich set of variations as a function of energy, pointing direction, orbital phase and total time in orbit. Since the typical OSSE signal-to-noise ratio is very low, understanding and accounting for background variations is a critical and complex task.

In order to deal with the OSSE data analysis challenge, a system of databases and software have been developed. These include a series of well defined data products that are handled first by a production data analysis system performing routine processing and then by a scientific analysis system which allows individual scientists to examine and manipulate OSSE data.

II. OSSE Data Products

The primary OSSE data products are described in Table 1. These are divided into "Level 1" data products and "Level 2" data products. The former have no instrument response deconvolution or data selection applied while the latter are more directly interpretable, although instrument bias may remain in some cases. Table 2 indicates the amount of data produced during a typical two-week observation period.

Table 1 -- OSSE Data Products

1. Level 1 Data Products

- 2-minute Spectra -- These spectra are summed from raw OSSE spectra at the basic 4-32 second integration interval. They include extensive header information with instrument configuration and rate data and orbital ephemerides and environment. Primary detector spectra and calibration data but no diagnostic or shield data are typically included.

- Background-Subtracted 2-minute Spectra -- Similar to the raw 2-minute spectra but with a standard background subtraction algorithm applied. To achieve the highest sensitivities, modified background subtraction algorithms may need to be applied. Propagated uncertainties are included but background spectra are not.

- Pulsar Data -- Event-by-event data with time resolution down to 0.125 msec but with limited energy band coverage and rate data with time resolution down to 4 msec are available, depending on instrument mode.

- Burst Data -- 4096 time bins measured from shields. Time resolution is adjustable down to 4 msec resolution.

2. Level 2 Data Products

- Integrated Count Spectra -- Background-subtracted spectra summed over some or all of one or more two-week observing periods. Data selection criteria have been applied to reduce systematic effects.

- Estimated Incident Photon Spectra -- These are the result of instrument response deconvolution being applied to integrated count spectra. Depending on the algorithm used, these may or may not be model dependent.

- Sky Maps -- Estimated celestial photon flux in a given energy band as a function of position on the sky. These will only be available for portions of the sky where mapping programs were undertaken.

- Integrated Pulsar Light Curves -- Epoch-folded pulsar light curves integrated over selected portions of an observation. Spectra are available for each phase bin. These are only available for periods when pulsar mode is active.

3. Auxiliary Data

- Instrument Model -- Describes instrument response vs. energy, incident angle, instrument configuration, and temperature. Subsets or reduced versions may be adequate for most analyses.

- Instrument State History -- Instrument configuration vs. time.

Table 2 -- Data Volumes

Product	Volume
2-minute Spectra	~260 Mbytes/2 wk. observation
Bkgd-subtracted 2-minute Spectra	~110 Mbytes/2 wk. observation
Pulsar Data (typical, where present)	~480 Mbytes/2 wk. observation
Instrument Model	~100 Mbytes total
Standard Plots (screening plots)	~70 Mbytes/2 wk. observation
Total Data	~920 Mbytes/2 wk. observation

OSSE data products will be transported using the FITS format with the FITS tables extension. The transport medium will, at least initially, be 6250 bpi 9-track magnetic tapes. The data volumes listed in table 2 would occupy approximately seven tapes.

III. OSSE Data Analysis

OSSE data analysis flow is illustrated in figures 1 - 5. OSSE telemetry data are distributed from PACOR at GSFC via 56 kbps telephone line to NRL and, as a backup, to Northwestern University (figure 1). Routine production analysis, as described in the following paragraphs, takes place at NRL. Archive media are distributed to the OSSE Co-I institutions as required in a timely fashion. Science analysis takes place primarily at NRL and Northwestern, although the other Co-I institutions are not precluded from participation as part of the PI team.

Production analysis (figure 2) consists of routine procedures performed on all data as they are received. These include reception, quality analysis, archiving, decommutation into separate data streams for various types of data, automatic summation into pointing intervals, database loading, automatic calibration and transient searching algorithm execution and hardcopy record production. Production analysis takes as input the telemetry stream as received from PACOR. It also has access to auxiliary data concerning orbital environment (magnetic field, SAA) and solar system barycenter vector. These auxiliary data are combined with the incoming packets to form messages of augmented packets which are written to magnetic disk.

A second pass through the data is used to decommutate the individual packets and form functional data streams summed into intervals that are usually synchronized to OSSE pointing intervals (typically 2 minutes). The data streams formed include detector spectra, calibration spectra, diagnostic spectra (e.g. shield spectra, charged particle detector spectra, individual PMT spectra, etc.), burst time histories, and instrument model data (e.g. pulse height analyzer calibrations). These data streams form the basis for most scientific analyses.

In addition five auxiliary processes are run during production analysis. These are: 1) automatic calibration spectrum line fitting; 2) automatic transient searching; 3) production of standardized data screening plots, both in electronic and film formats; 4) production of an observation history database containing entries for each source observation period (typically two entries per orbit); and 5) production of an instrument state history detailing changes in the commanded state of the instrument.

The data are archived to several distinct archival media for ease of access. Augmented packets are archived to one media (initially, magnetic tape), while various decommutated data streams are archived to separate media. Archiving to optical disk is under consideration for future revisions.

Production analysis functions are routinely performed at NRL within 24 hours of data reception. Archive and derived media are shipped to Northwestern University on a regular and frequent basis. Production analysis procedures are largely transparent to the scientific user. A limited number of applications may require either special production runs, in particular if spectra on time scales faster than the 2-minute pointing interval are desired, or specialized telemetry archive analysis routines. These routines, beyond a basic set, are supplied on a need basis.

OSSE data is routinely screened by data technicians looking for telemetry problems, glitches in housekeeping information, orbital environment information, and time sequence, instrument performance problems, etc. When problems are found, data corrections are entered into an on-line database of archive modifications.

The archived decommutated data streams form the basis for most scientific analysis. These may be accessed by the individual scientific user. A tape and data file database system allows the user to select the volumes for the interval, observed source, etc. of interest. A production-science interface program (figure 3) reads the archive media, searches the archive modification database for changes to incorporate, and builds a disk-based database for the individual user. In addition, the user has access to on line databases containing instrument state information and the instrument model as produced by the production analysis process and the OSSE calibration analysis effort.

Analysis of pulsar data represents a special case. Since these data are not easily compressed, no separate pulsar data stream is formed during production processing. Rather, the production science interface program can read specified (via reference to the observation history) intervals of pulsar data directly from the augmented packet archive media and produce disk-based or tape-based pulsar data streams for the observation periods of interest.

Scientific analysis consists of processes required for the scientific user to interact with the data with the goal of producing a scientifically usable result (e.g. estimates of production spectra from a source, models of source emission, etc.). Scientific analysis starts with semi-routine operations to remove instrumental effects. Foremost of these is background subtraction. Since all subsequent results are sensitive to errors in this procedure, it is included in the scientific rather than production analysis in order to give the scientific user more direct control over the process and to allow the user a feel for the compromises involved. As experience with OSSE data grows, background subtraction will become progressively more routine. However, certain situations (e.g. data near the limb of the earth), may always require special care.

Other scientific analysis procedures include instrument response unfolding and examination of results for systematic effects and artifacts. The latter will be aided by the ability to select data sets from the spectral database while applying a wide variety of instrument state and environmental criteria. These procedures will include specialized algorithms for dealing with such situations as source confusion and spatially extended source analysis.

At least until a large base of experience in dealing with OSSE data is acquired, data selection and review of results will require significant expertise. Evaluation of systematic effects and uncertainties as a function of instrument parameters and observation parameters (such as earth angle) is a critical task that is not easy to automate. A large part of OSSE analysis consists of repeated selections of data for different ranges of parameters, weighing statistics versus systematics. Once results are obtained, verifying the reality and significance of those results will demand a deep knowledge of OSSE background and other systematic effects. Experience with the SMM gamma ray spectrometer database has indicated that the variety of background-related artifacts is extremely rich and that artifacts vary on a wide range of time scales.

Instrument response deconvolution presents a special problem in that the response matrix is numerically ill-conditioned and hence cannot be inverted using conventional techniques. This is a well-known problem for gamma ray detectors in general and for scintillation detectors specifically. The traditional approach to this problem in astrophysical applications has been to use an incident photon flux model with a finite (and usually small) number of free parameters. The model is folded through the instrument response then compared to the observed pulse height analyzer spectrum. Parameters are adjusted to achieve a best fit. A deconvolved spectrum can then, for example, be calculated by comparing the variations of the actual PHA spectrum from the modeled PHA spectrum, then applying these variations to the model incident photon spectrum. This results in a rather model dependent estimated incident photon spectrum, in addition to other problems. Hence true level 1 data, with instrument effects removed but without any interpretation applied, cannot be obtained using this technique. Any scientist wishing to try his own models against the data must become familiar with the deconvolution process and instrument model. This will not become routine until the OSSE P.I. team has had a considerable amount of experience in assessing the systematic uncertainties associated with this process in a flight environment.

Alternative deconvolution techniques are available (e.g. maximum entropy, Backus-Gilbert, etc.). However, these have not been used routinely in the past for gamma ray astronomy applications and time will be required to gain confidence in their results.

Final results are derived by a variety of processes such as combination and summation of spectra in a statistically meaningful manner, combining results from different pointing positions into sky maps, modeling spectra and spatial distributions and performing time-series analysis. Presentation and publication quality graphics utilities will be available to display results.

Scientific data analysis, resulting in interpretable (level 2) data products, is conducted using applications attached to a user shell or environment. This is referred to as the Integrated GRO/OSSE Reduction Environment (IGORE; see figure 4). IGORE is based on IDL, a commercially available data handling language. IGORE consists of the following components: 1) the IDL programmable Higher Level Command Language (HLCL) modified to handle OSSE data structures in a convenient, integrated fashion; 2) an interface protocol between IDL and FORTRAN applications (known as the Application Front End or AFE); 3) database query and update applications; 4) data i/o applications; 5) data analysis applications; and 6) a custom error and condition handling system.

OSSE databases generally consist of data in an indexed OSSE format and a relational database of data descriptors describing and pointing to the data. Spectra, pulsar data, time histories and maps are all stored in this fashion. The INGRES relational database management system is used to store the descriptor databases. Data selection tasks are normally accomplished by applications using INGRES forms to access the descriptor databases.

Since each analyst builds his own database tables for his use while analyzing the data, he may modify and add to these tables at will as he creates intermediate derived results. On completion of a given project, all derived results are reviewed by the OSSE project and, on acceptance, incorporated into the archival OSSE database. An analysis history database is maintained as part of the observation history and contains pointers to these data.

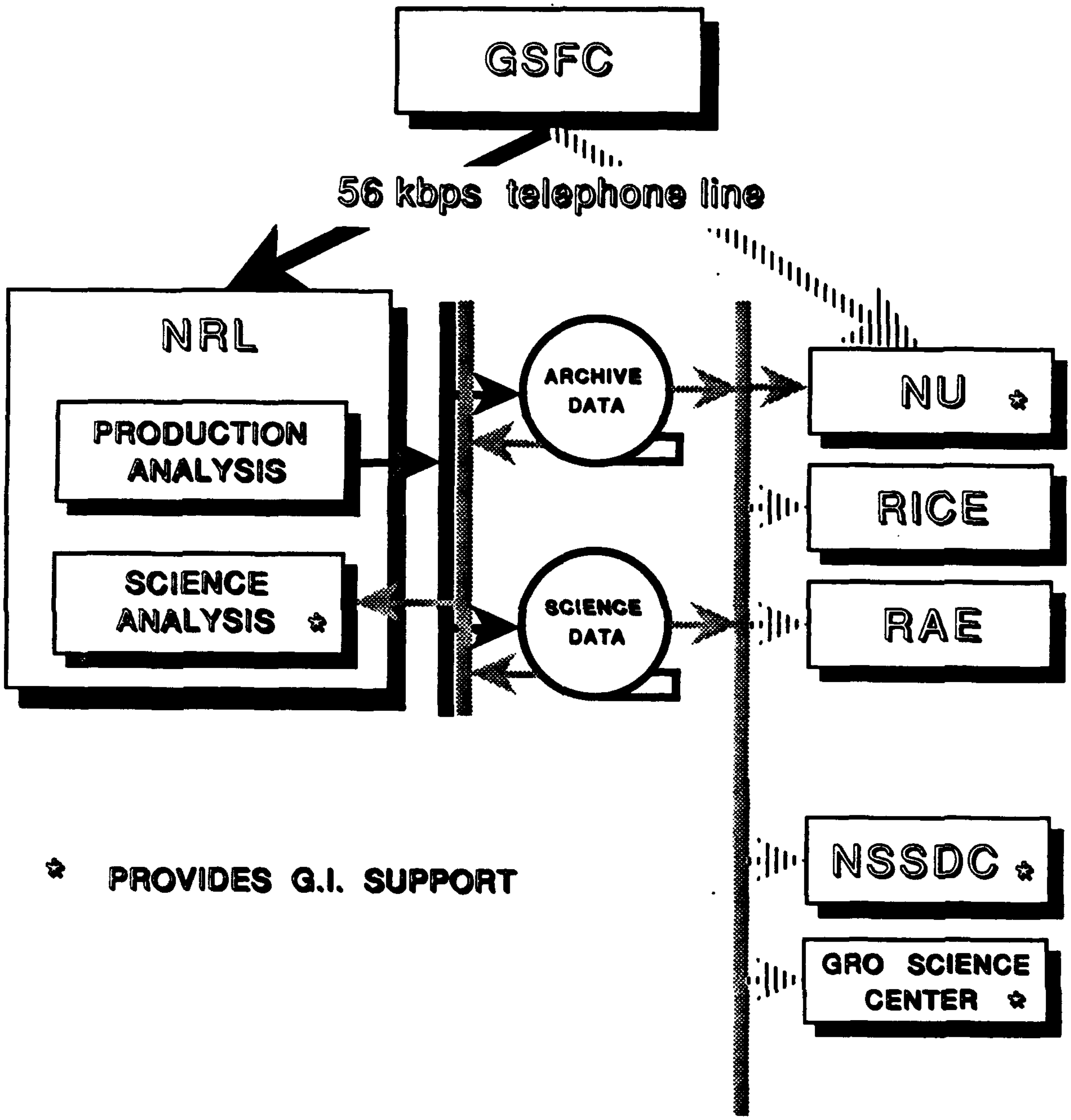
Available analysis applications are indicated in figure 5. User-written applications may also be integrated into the system. Applications that meet testing and documentation standards may be configuration managed and linked into the standard IGORE system. Other applications may be dynamically linked at run time using the IDL LINKIMAGE command. The AFE for any application is generated by a preprocessor that "reads" information embedded in the application source code and "writes" the AFE source.

IV. Conclusions

OSSE data analysis is a complex enterprise due to a variety of data types and instrument operating modes and due to the low signal-to-noise operating regime. In order to achieve ultimate sensitivity, a great deal of insight into the functioning of the instrument and its interactions with the orbital environment will be required. A system of data and software systems has been developed to deal with OSSE data, but the scientist must expect to interact with the data analysis process at various fundamental levels.

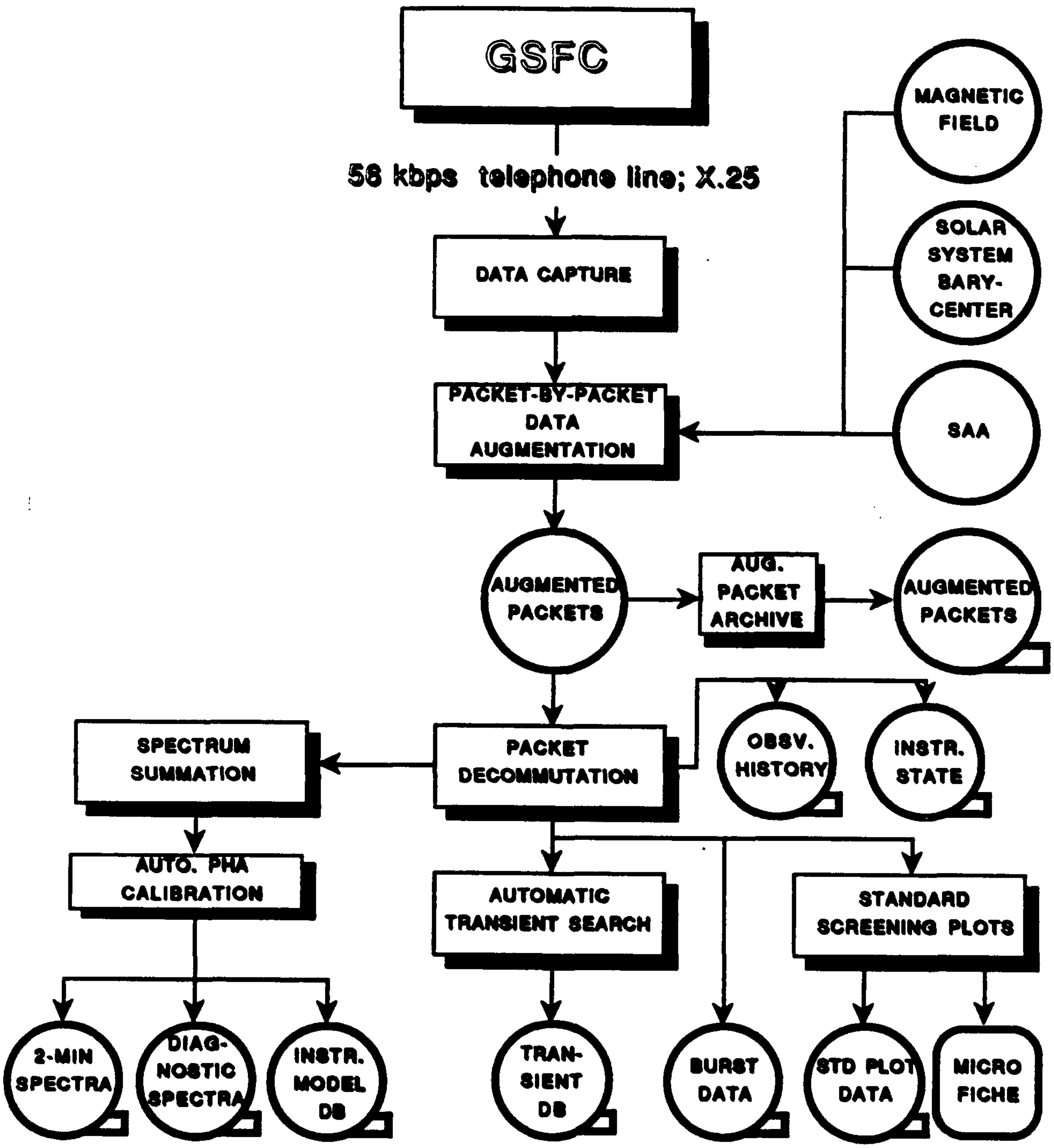
OSSE DATA FLOW

FIG 1: INSTITUTIONAL



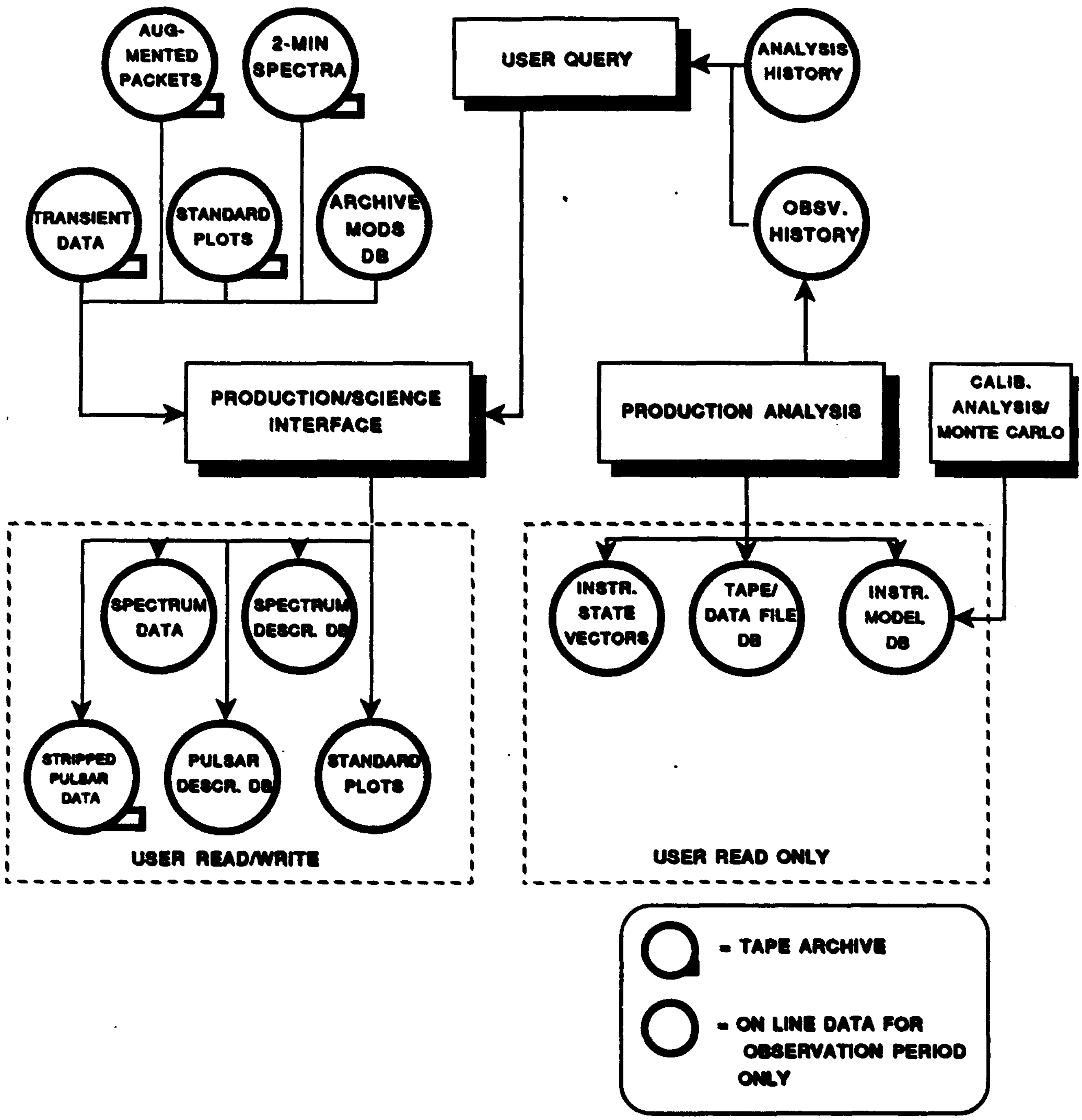
OSSE DATA FLOW

FIG 2: PRODUCTION ANALYSIS



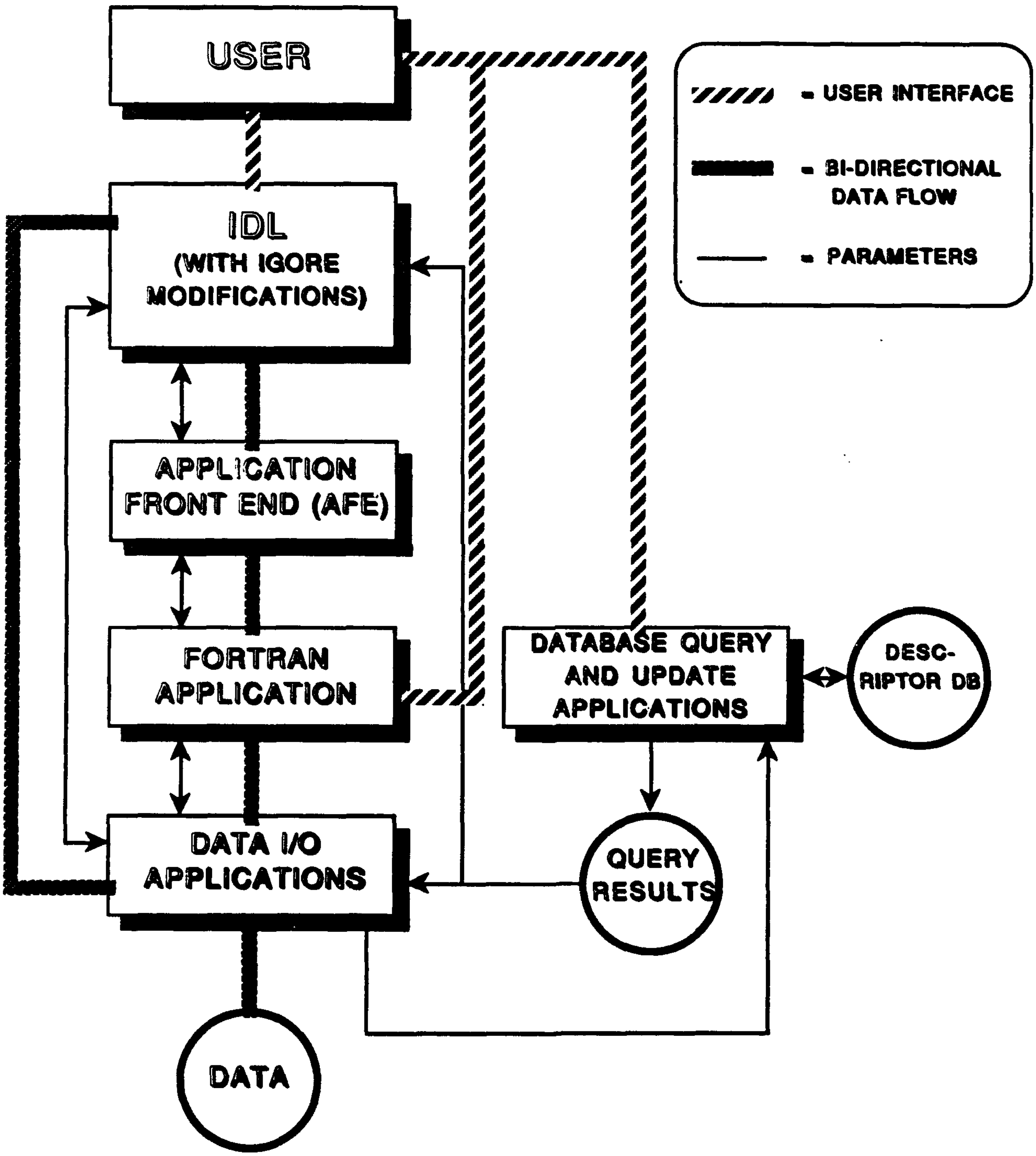
OSSE DATA FLOW

FIG 3: INTEGRATED DATABASE SYSTEM



OSSE DATA FLOW

FIG 4: INTEGRATED GRO/OSSE REDUCTION ENVIRONMENT (IGORE)



OSSE DATA FLOW

FIG 5: MAJOR IGORE APPLICATIONS